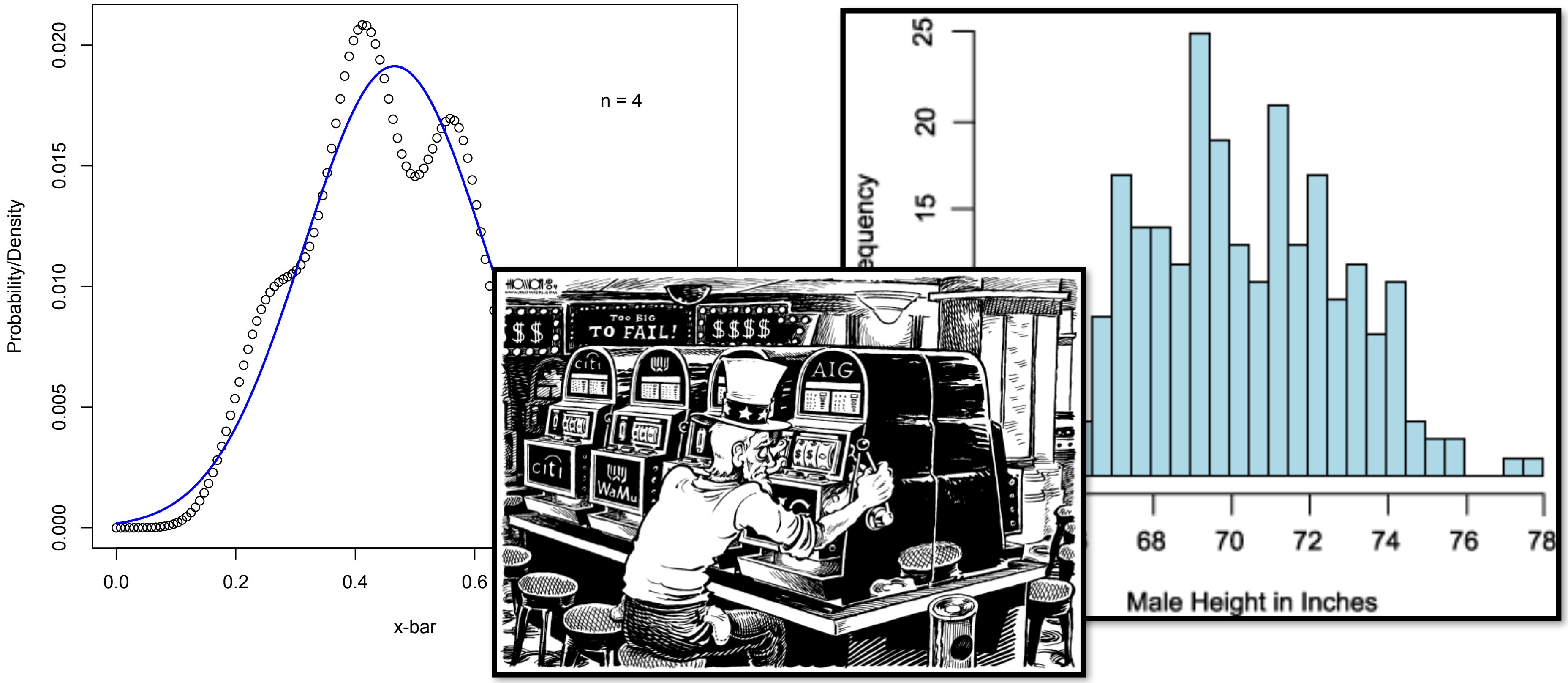


$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)=\mu\right)=1$$

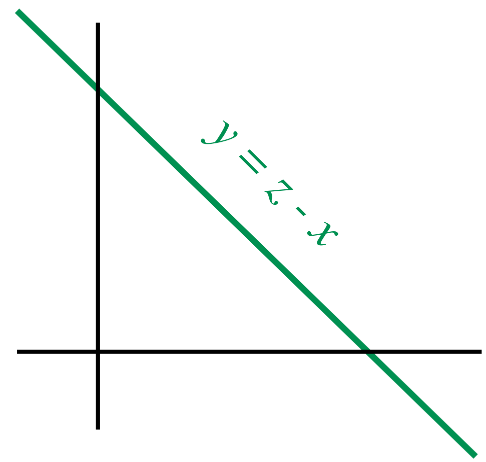
If X, Y are independent, what is the distribution of Z = X + Y ?

Discrete case:

$$p_Z(z) = \Sigma_x\, p_X(x) \bullet p_Y(z-x)$$

Continuous case:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \bullet f_Y(z-x)\, dx$$

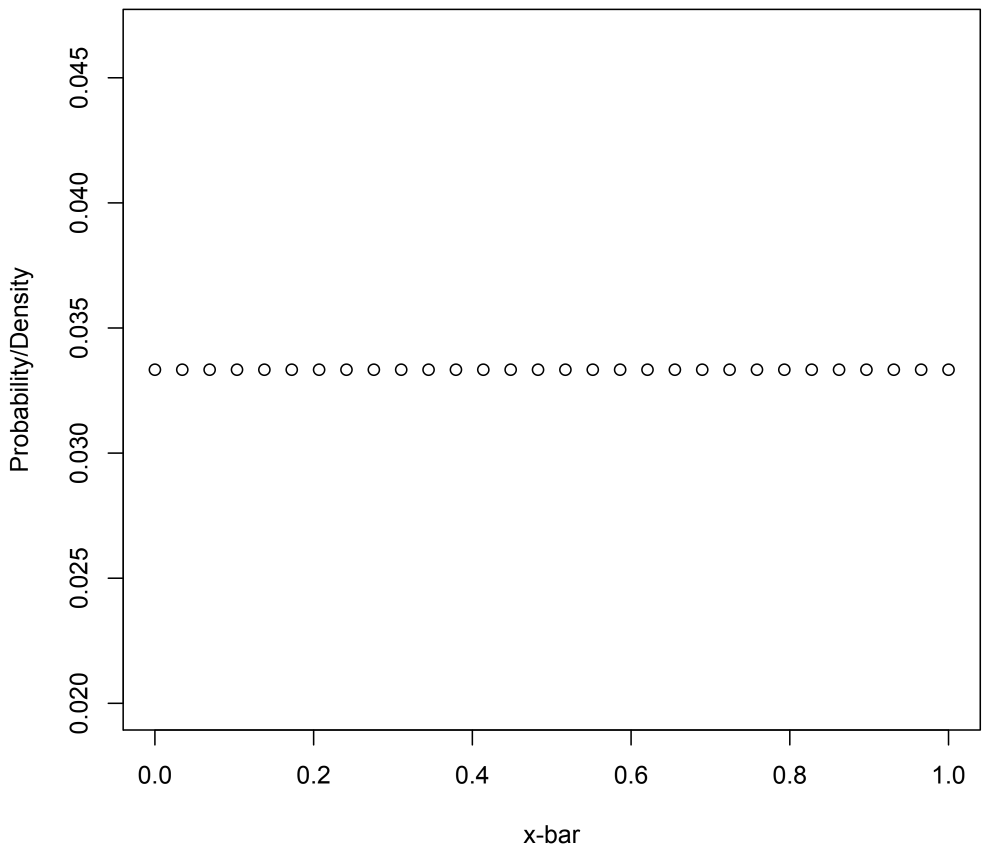


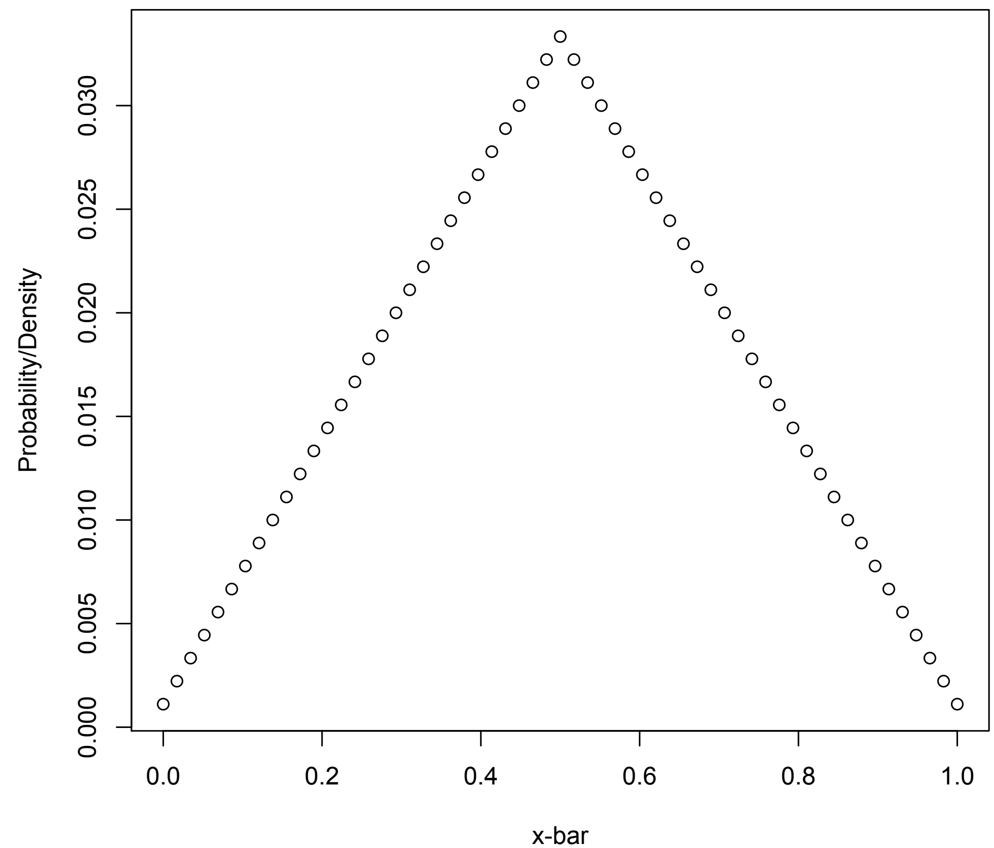
E.g. what is the p.d.f. of the sum of 2 normal RV's?

W = X + Y + Z ? Similar, but double sums/integrals

V = W + X + Y + Z ? Similar, but triple sums/integrals

If X and Y are *uniform*, then Z = X + Y is *not*; it's *triangular* (like dice):





Intuition: X + Y ≈ 0 or ≈ 1 is rare, but many ways to get X + Y ≈ 0.5

aka transforms; b&t 229

Powerful math tricks for dealing with distributions

We won't do much with it, but mentioned/used in book, so a very brief introduction:

The k^{th} moment of r.v. X is $E[X^k]$; M.G.F. is $M(t) = E[e^{tX}]$

$$e^{tX} = X^0\frac{t^0}{0!} + X^1\frac{t^1}{1!} + X^2\frac{t^2}{2!} + X^3\frac{t^3}{3!} + \cdots$$

$$M(t) = E[e^{tX}] = E[X^0]\frac{t^0}{0!} + E[X^1]\frac{t^1}{1!} + E[X^2]\frac{t^2}{2!} + E[X^3]\frac{t^3}{3!} + \cdots$$

$$\frac{d}{dt}M(t) = 0 + E[X^1] + E[X^2]\frac{t^1}{1!} + E[X^3]\frac{t^2}{2!} + \cdots$$

$$\frac{d^2}{dt^2}M(t) = 0 + 0 + E[X^2] + E[X^3]\frac{t^1}{1!} + \cdots$$

$$\frac{d}{dt}M(t)\Big|_{t=0} = E[X] \qquad \frac{d^2}{dt^2}M(t)\Big|_{t=0} = E[X^2] \quad \ldots \quad \frac{d^k}{dt^k}M(t)\Big|_{t=0} = E[X^k] \quad \ldots$$

Closely related to Laplace transforms, which you may have seen.

An example:

MGF of normal(μ,σ^2) is $\exp(\mu t+\sigma^2 t^2/2)$

Two key properties:

1. MGF of *sum* independent r.v.s is *product* of MGFs:

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}\, e^{tY}] = E[e^{tX}]\, E[e^{tY}] = M_X(t)\, M_Y(t)$$

2. Invertibility: MGF uniquely determines the distribution.

e.g.: $M_X(t) = \exp(at+bt^2)$, with $b>0$, then $X \sim \text{Normal}(a,2b)$

Important example: *sum of independent normals is normal:*

$$X\sim\text{Normal}(\mu_1,\sigma_1^2) \quad Y\sim\text{Normal}(\mu_2,\sigma_2^2)$$

$$M_{X+Y}(t) = \exp(\mu_1 t + \sigma_1^2 t^2/2) \cdot \exp(\mu_2 t + \sigma_2^2 t^2/2)$$

$$= \exp[(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2]$$

So X+Y has mean $(\mu_1+\mu_2)$, variance $(\sigma_1^2+\sigma_2^2)$ (duh) *and is normal!*

(*way* easier than slide 2 way!)

Consider i.i.d. (independent, identically distributed) R.V.s

$X_1, X_2, X_3, \ldots$

Suppose X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i] < \infty$. What are the mean & variance of their sum?

$$E[\sum_{i=1}^{n} X_i] = n\mu \text{ and } Var[\sum_{i=1}^{n} X_i] = n\sigma^2$$

So limit as $n \to \infty$ *does not exist* (except in the degenerate case where $\mu = 0$; note that if $\mu = 0$, the *center* of the data stays fixed, but if $\sigma^2 > 0$, then the variance is unbounded, i.e., its *spread* grows with n).

Consider i.i.d. (independent, identically distributed) R.V.s

$X_1, X_2, X_3, \ldots$

Suppose X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i] < \infty$

What about the *sample mean* $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, as $n \to \infty$?

$$\mathrm{E}\left[\, M_n \right] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

$$\mathrm{Var}\left[\, M_n \right] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

So, limits *do* exist; mean is independent of *n*, variance shrinks.

Continuing: iid RVs $X_1, X_2, X_3, \ldots$; $\mu = E[X_i]$; $\sigma^2 = Var[X_i]$; $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathrm{E}\left[\, M_n \right] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \text{ ;} \mathrm{Var}\left[\, M_n \right] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

Expectation is an important guarantee.

BUT: observed values may be far from expected values.

E.g., if X_i ~ Bernoulli(½), the E[X_i]= ½, but X_i is *NEVER* ½.

Is it also possible that sample mean of X_i's will be far from ½?

Always? Usually? Sometimes? Never?

b&t 5.2

For any ε > 0, as n → ∞

$$\Pr\left(|M_n - \mu| > \epsilon\right) \to 0$$

Proof: (*assume* $\sigma^2 < \infty$; theorem true without that, but harder proof)

$$\mathrm{E}\left[\, M_n \right] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

$$\mathrm{Var}\left[\, M_n \right] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

By Chebyshev inequality,

$$\Pr\left(|M_n - \mu| > \epsilon\right) \le \frac{\sigma^2}{n\epsilon^2} \quad \xrightarrow{n \to \infty} \quad 0$$

b&t 5.5

i.i.d. (independent, identically distributed) random vars

$X_1, X_2, X_3, \ldots$

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

X_i has $\mu = E[X_i] < \infty$

$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mu\right) = 1$$

Strong Law ⇒ Weak Law (but not vice versa)

Strong law implies that for any ε > 0, there are only a finite number of n satisfying the weak law condition $|M_n - \mu| \geq \epsilon$ (almost surely, i.e., with probability 1)

Supports the intuition of probability as long term frequency

Weak Law:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Strong Law:

$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mu\right) = 1$$

How do they differ? Imagine an infinite 2-D table, whose rows are indp infinite sample sequences X_i. Pick ε. Imagine cell m,n lights up if average of 1^{st} n samples in row m is > ε away from μ.

WLLN says fraction of lights in n^{th} column goes to zero as n →∞. It does not prohibit every row from having ∞ lights, so long as frequency declines.

SLLN also says only a vanishingly small fraction of rows can have ∞ lights.

weak vs strong laws – supplement

The differences between the WLLN & SLLN are subtle, and not critically important for this course, but for students wanting to know more (e.g., not on exams), here is my summary. Both "laws" rigorously connect long-term averages of repeated, independent observations to mathematical expectation, justifying the intuitive motivation for E[.]. Specifically, both say that the sequence of (non-i.i.d.) rvs $M_n = \sum_{i=1}^{n} X_i/n$ derived from any sequence of i.i.d. rvs X_i converge to $E[X_i]=\mu$. The strong law totally subsumes the weak law, but the later remains interesting because (a) of its simple proof (Khintchine, early 20th century; using Cheybeshev's inequality (1867)) and (b) historically (WLLN was proved by Bernoulli ~1705, for Bernoulli rvs, >150 years before Chebyshev [Ross, p391]). The technical difference between WLLN and SLLN is in the definition of convergence.

<u>Definition</u>: Let Y_i be any sequence of rvs (i.i.d. not assumed) and c a constant.
Y_i converges to c in probability if

$$\forall \epsilon > 0, \lim_{n \to \infty} \Pr\left(|Y_n - c| > \epsilon\right) = 0$$

Y_i converges to c with probability 1 if

$$\Pr\left(\lim_{n \to \infty} Y_n = c\right) = 1$$

The weak law is the statement that M_n converges *in probability* to μ; the strong law states it converges *with probability 1* to μ. The strong law subsumes the weak law since convergence with probability 1 implies convergence in probability for any sequence Y_i of rvs (B&T problem 5.5-15). B&T ex 5.15 illustrates the failure of the converse. A second counterexample is given on the following slide.

Example: Consider the sequence of rvs $Y_n \sim$ Ber(1/n)

Recall the definition of convergence in probability:

$$\forall\epsilon > 0,\ \lim_{n\to\infty} \Pr\left(|Y_n - c| > \epsilon\right) = 0$$

Then Y_n converges to c = 0 in probability since $\Pr(Y_n > \varepsilon) = 1/n$ for any $0 < \varepsilon < 1$, hence the limit as $n\to\infty$ is 0, satisfying the definition.

Recall that Y_n converges to c with probability 1 if

$$\Pr\left(\lim_{n\to\infty} Y_n = c\right) = 1$$

However, I claim that $\lim_{n\to\infty} Y_n$ does not exist, hence doesn't equal zero with probability 1. Why no limit? A 0/1 sequence will have a limit if and only if it is all 0 after some finite point (i.e., contains only a finite number of 1's) or vice versa. But the expected number of 0's & 1's in the sequence are both infinite; e.g.:

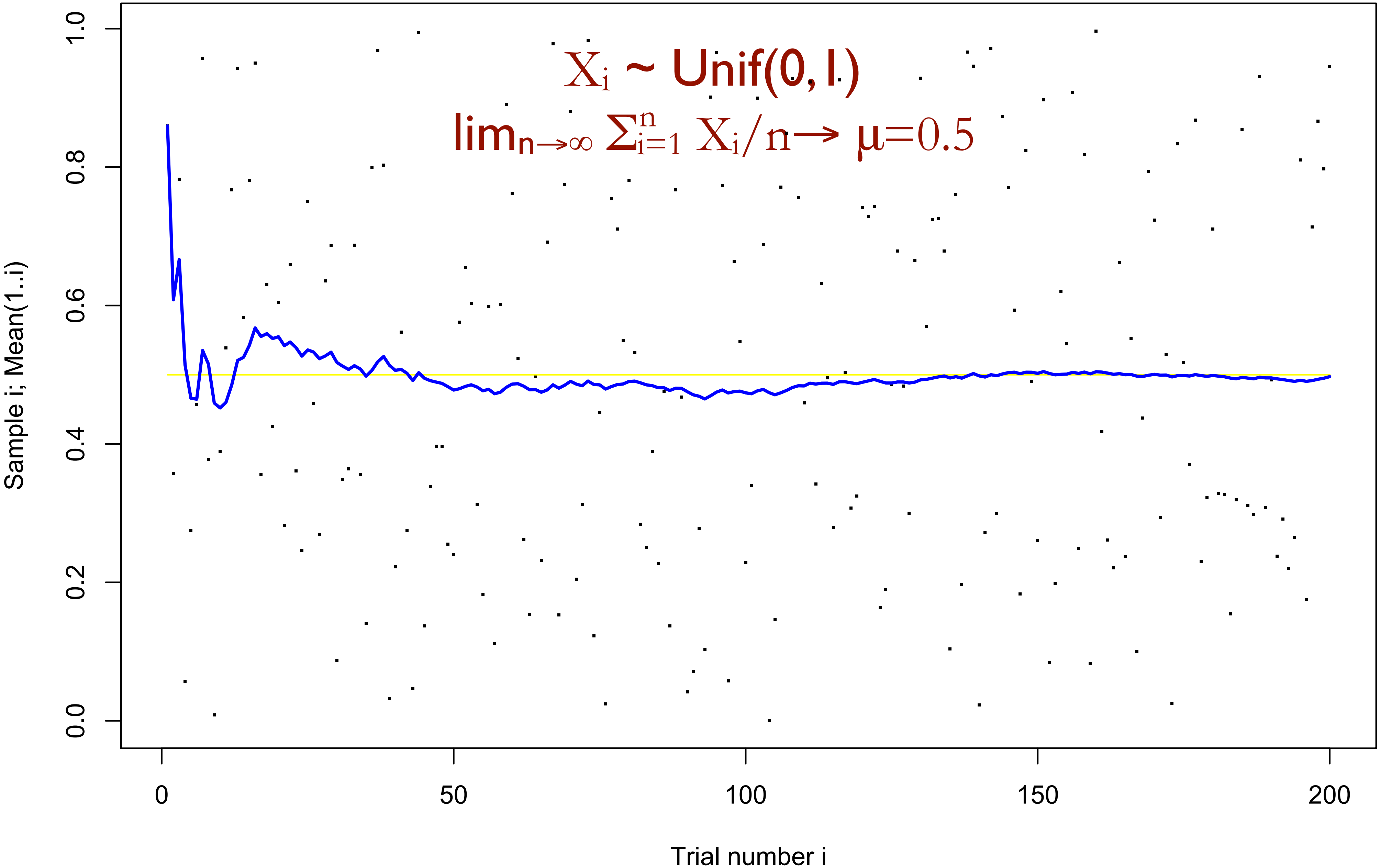
$$E\left[\sum_{i>0} Y_i\right] = \sum_{i>0} E[Y_i] = \sum_{i>0} \frac{1}{i} = \infty$$

Thus, Y_n converges in probability to zero, but does not converge with probability 1.

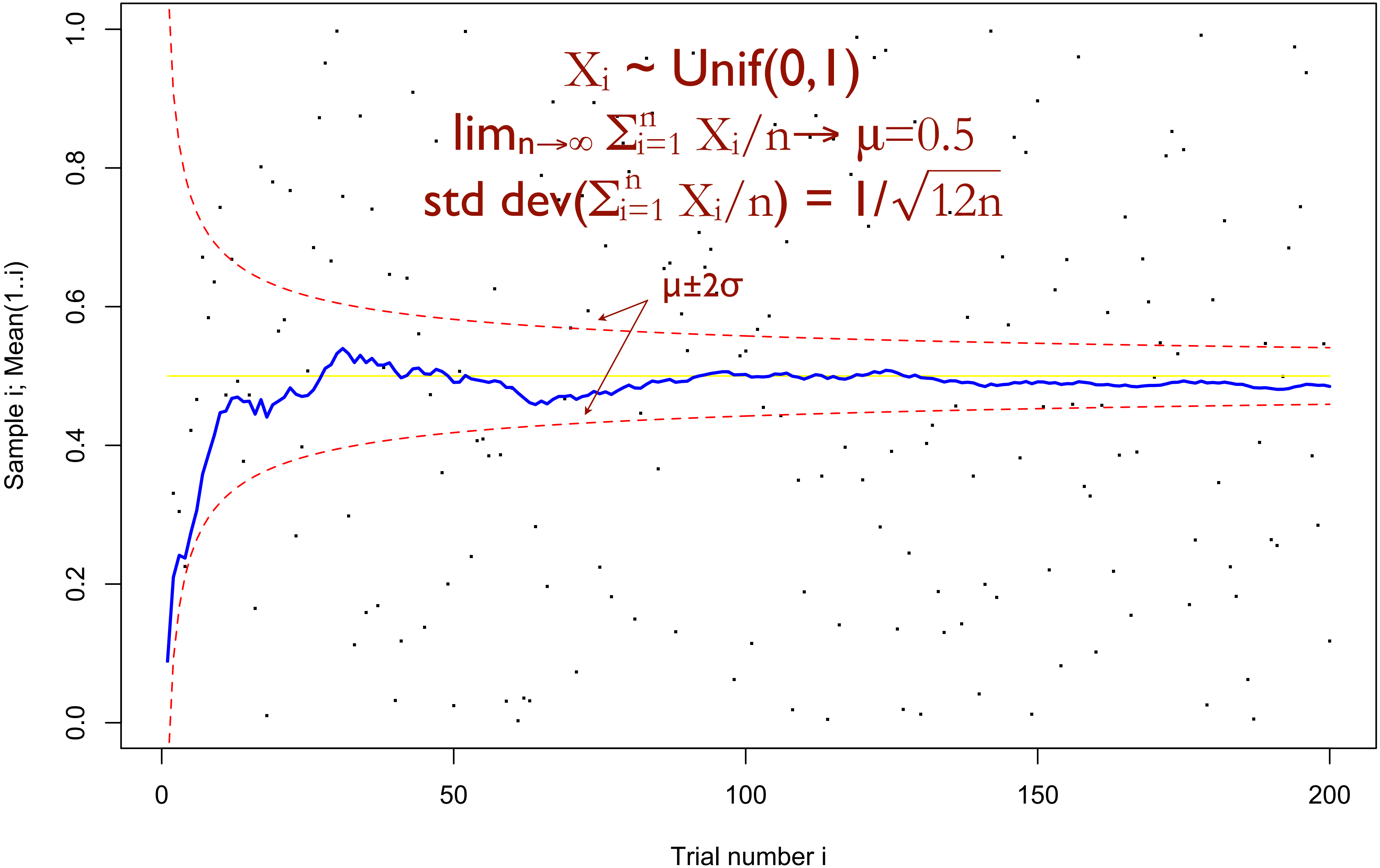
Revisiting the "lightbulb model" 2 slides up, w/ "lights on" ⇔ 1, column n has a decreasing fraction (1/n) of lit bulbs, while all but a vanishingly small fraction of rows have infinitely many lit bulbs.

(For an interesting contrast, consider the sequence of rvs $Z_n \sim$ Ber($1/n^2$).)

sample mean → population mean

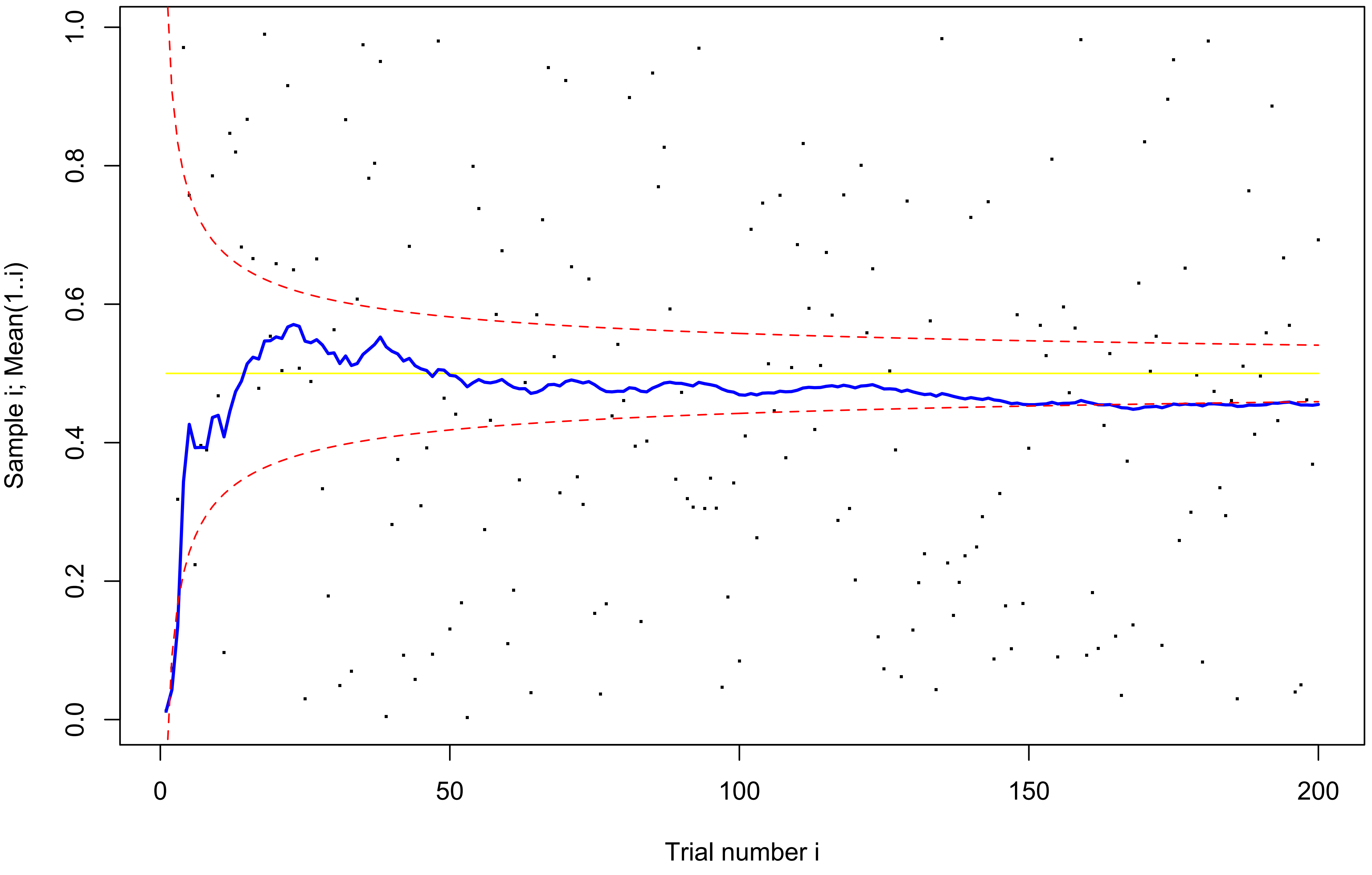


sample mean → population mean

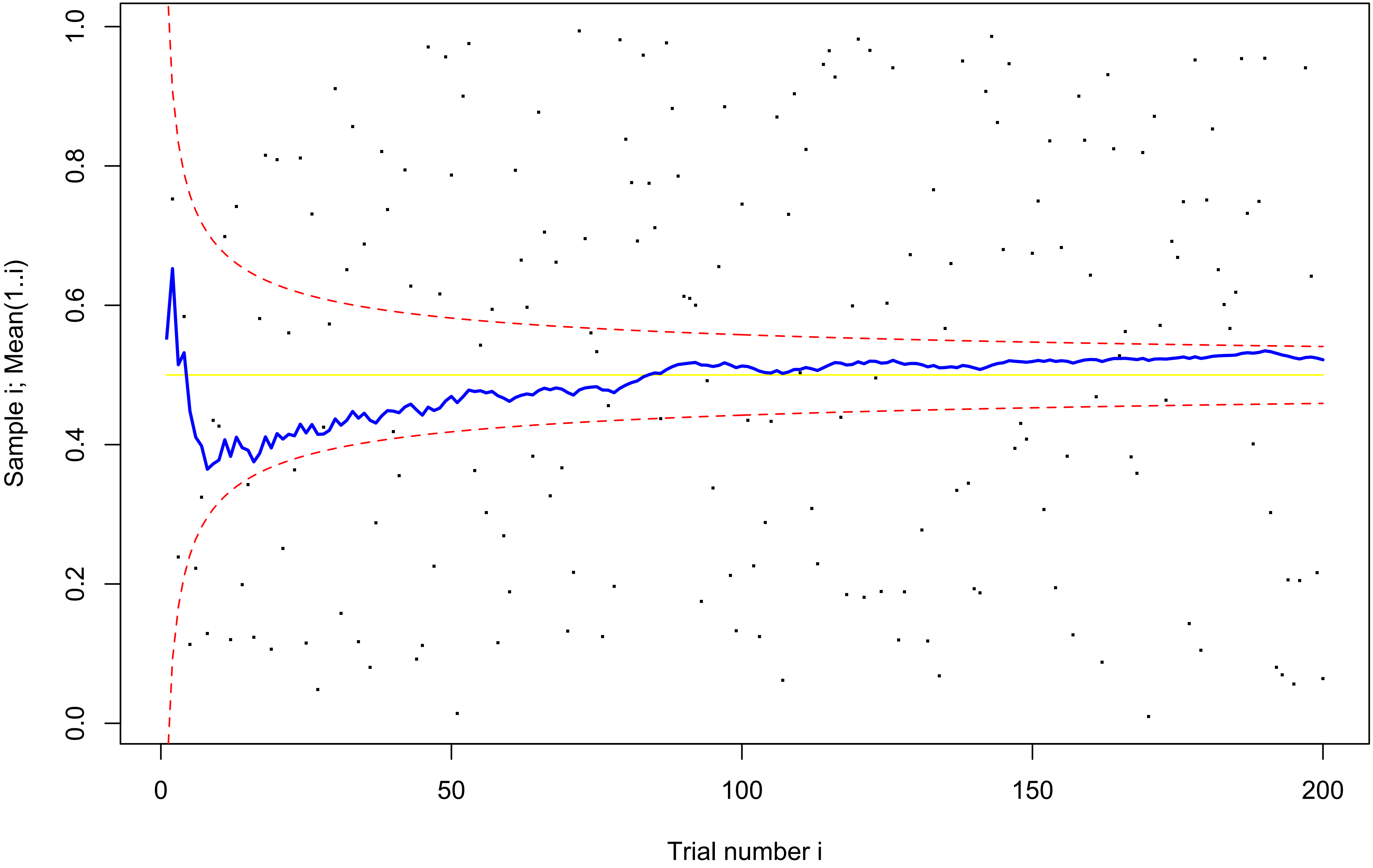


demo

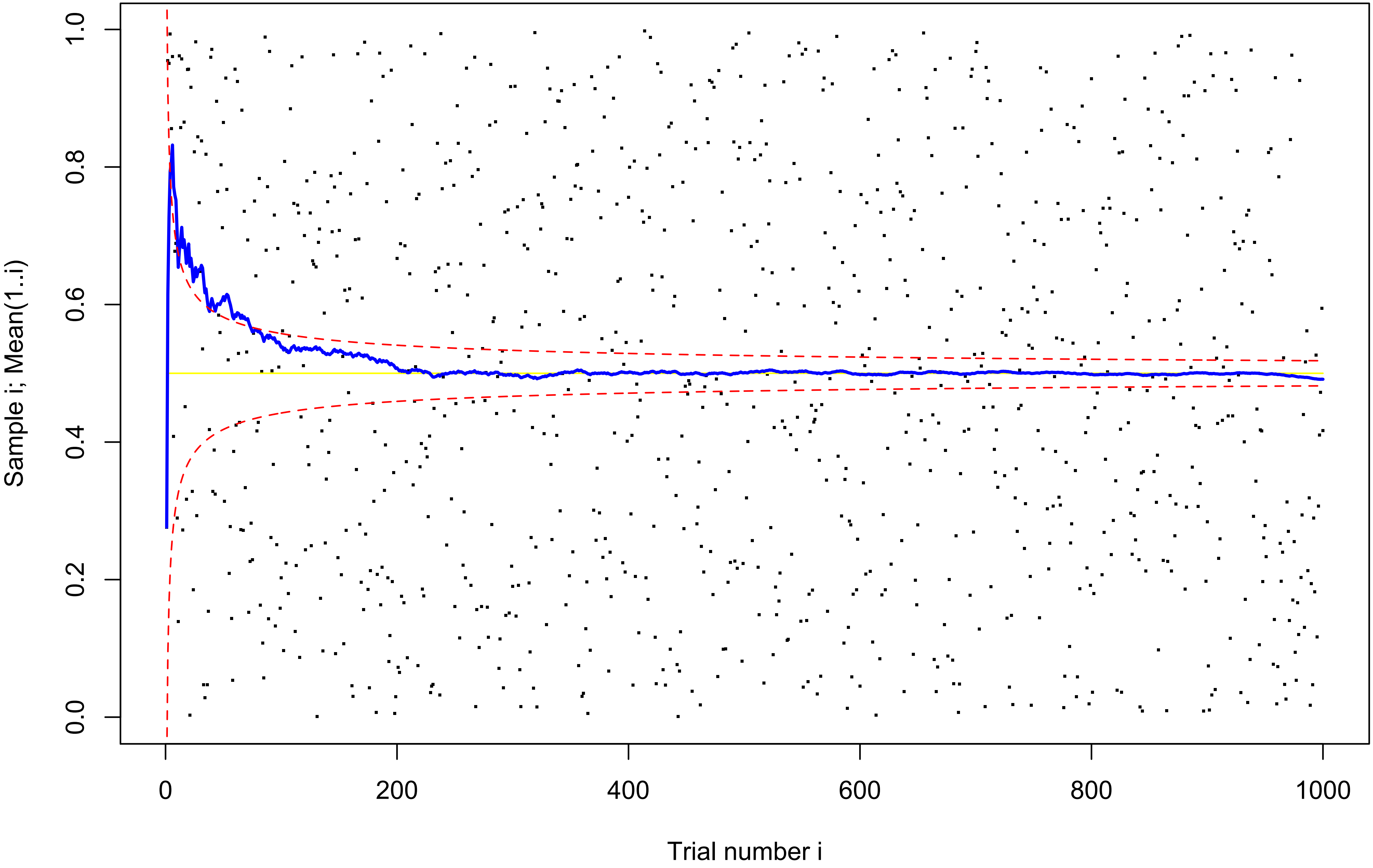
another example



another example



another example



Weak Law:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|>\epsilon\right)=0$$

Strong Law:

$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)=\mu\right)=1$$

How do they differ? Imagine an infinite 2-D table, whose rows are indp infinite sample sequences X_i. Pick ε. Imagine cell m,n lights up if average of 1st n samples in row m is > ε away from μ.

WLLN says fraction of lights in n^{th} column goes to zero as n →∞. It does not prohibit every row from having ∞ lights, so long as frequency declines.

SLLN also says only a vanishingly small fraction of rows can have ∞ lights.

Note: $D_n = E[\,|\,\Sigma_{1 \le i \le n}(X_i - \mu)\,|\,]$ *grows* with n, but $D_n/n \to 0$

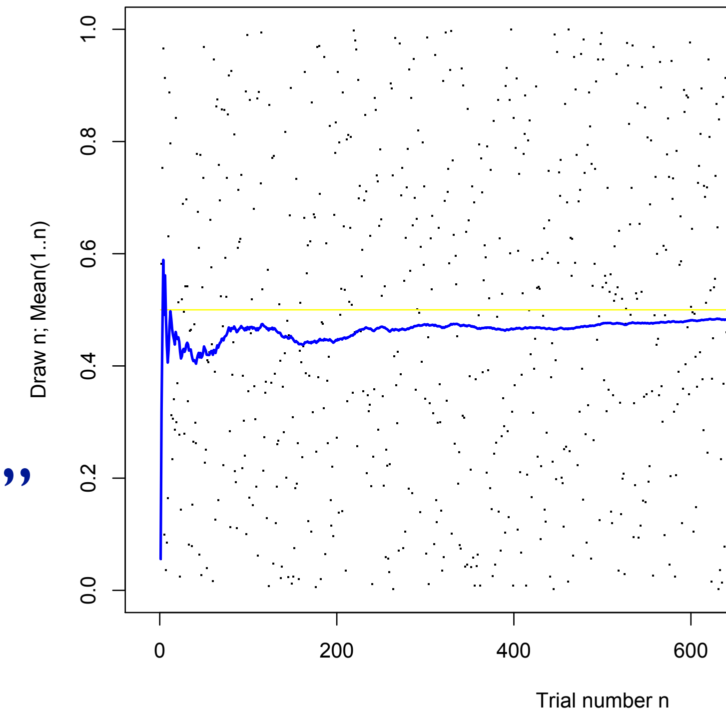
Justifies the "frequency" interpretation of probability

"Regression toward the mean"

Gambler's fallacy: "I'm *due* for a win!"

"Swamps, but does not compensate"

"Result will usually be close to the mean"



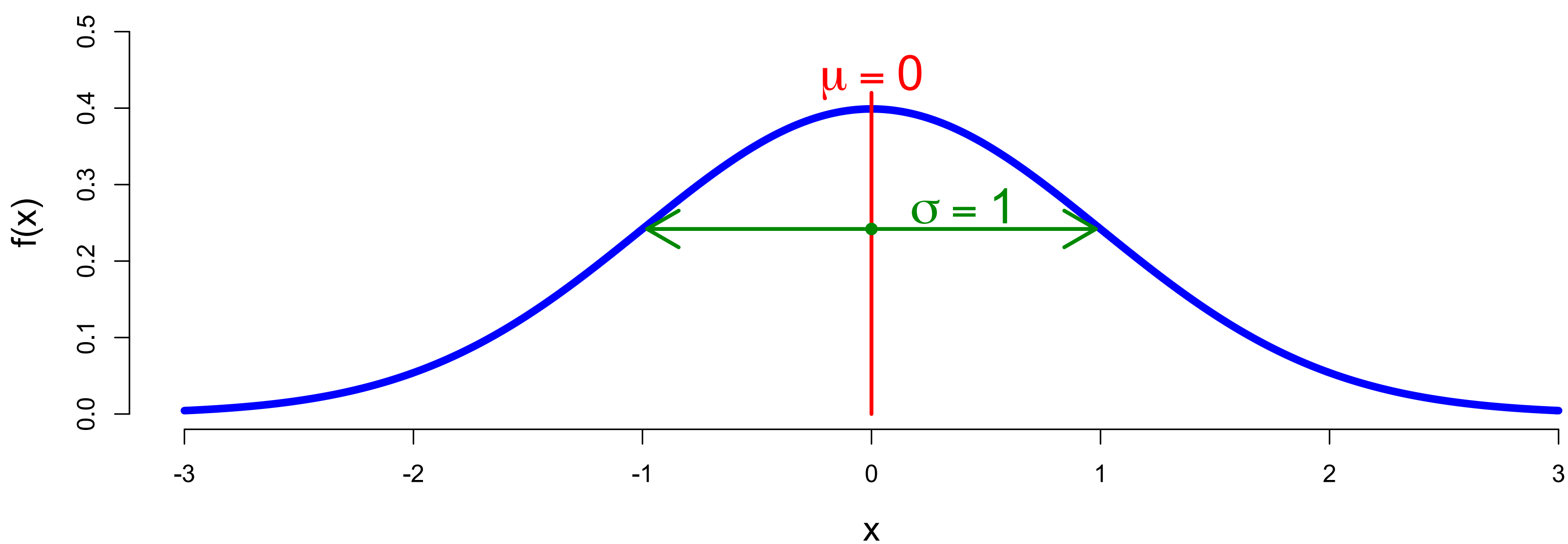
Many web demos, e.g.
http://stat-www.berkeley.edu/~stark/Java/Html/lln.htm

Recall

X is a normal random variable   X ~ N(μ,σ²)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$E[X] = \mu \qquad \mathrm{Var}[X] = \sigma^2$$



i.i.d. (independent, identically distributed) random vars

$X_1, X_2, X_3, \ldots$

X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i] < \infty$

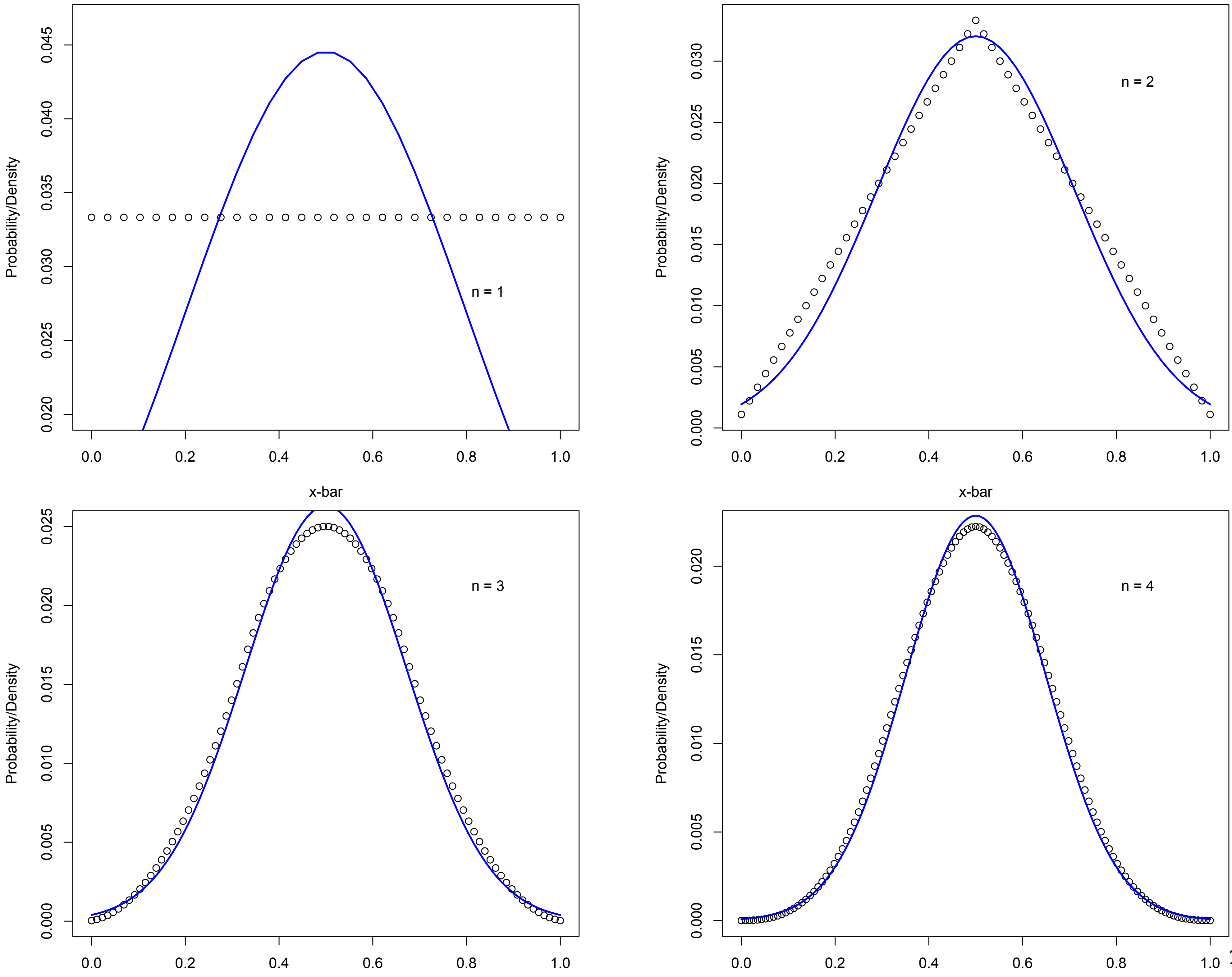
As $n \to \infty$,

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

Restated: As $n \to \infty$,

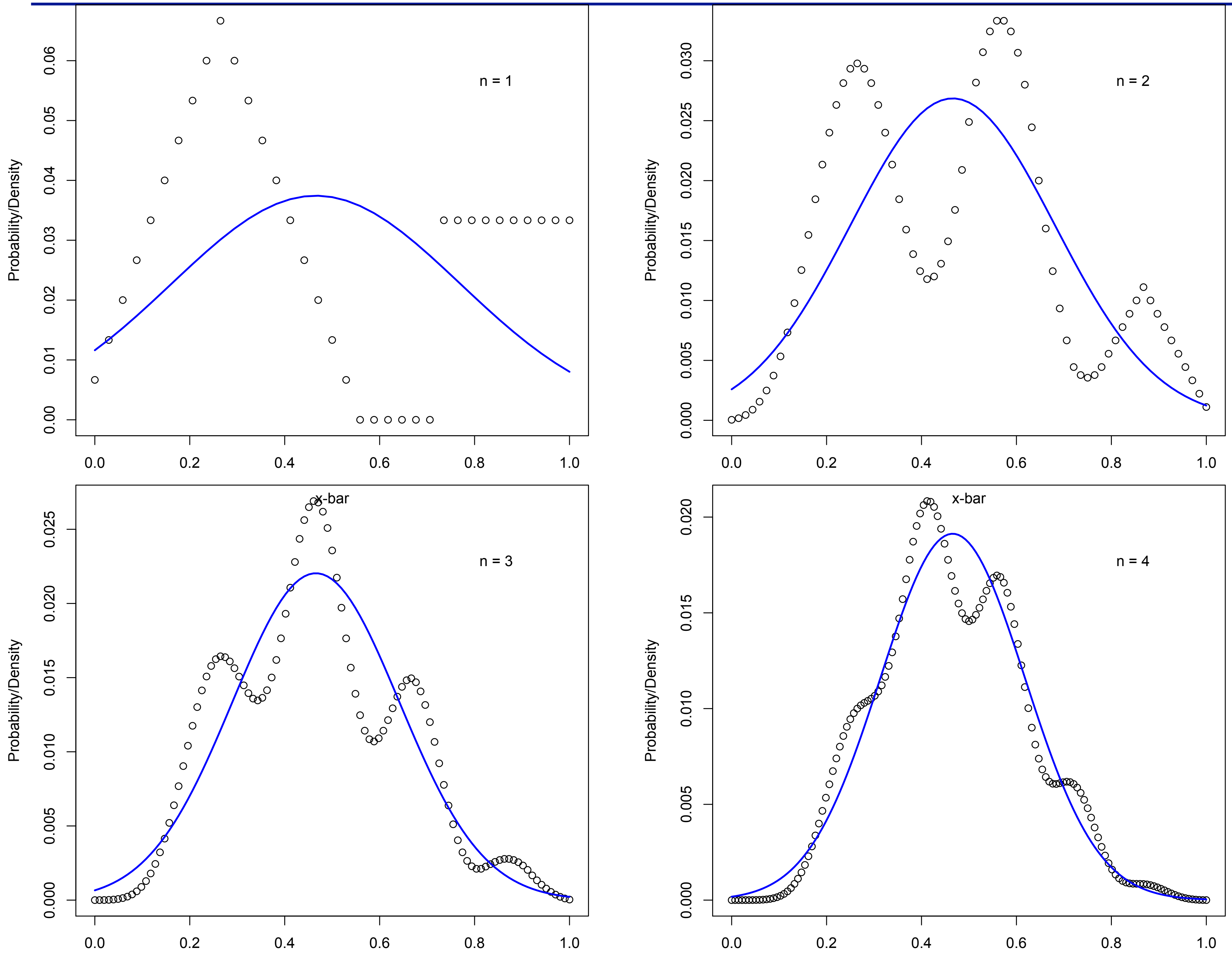
$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

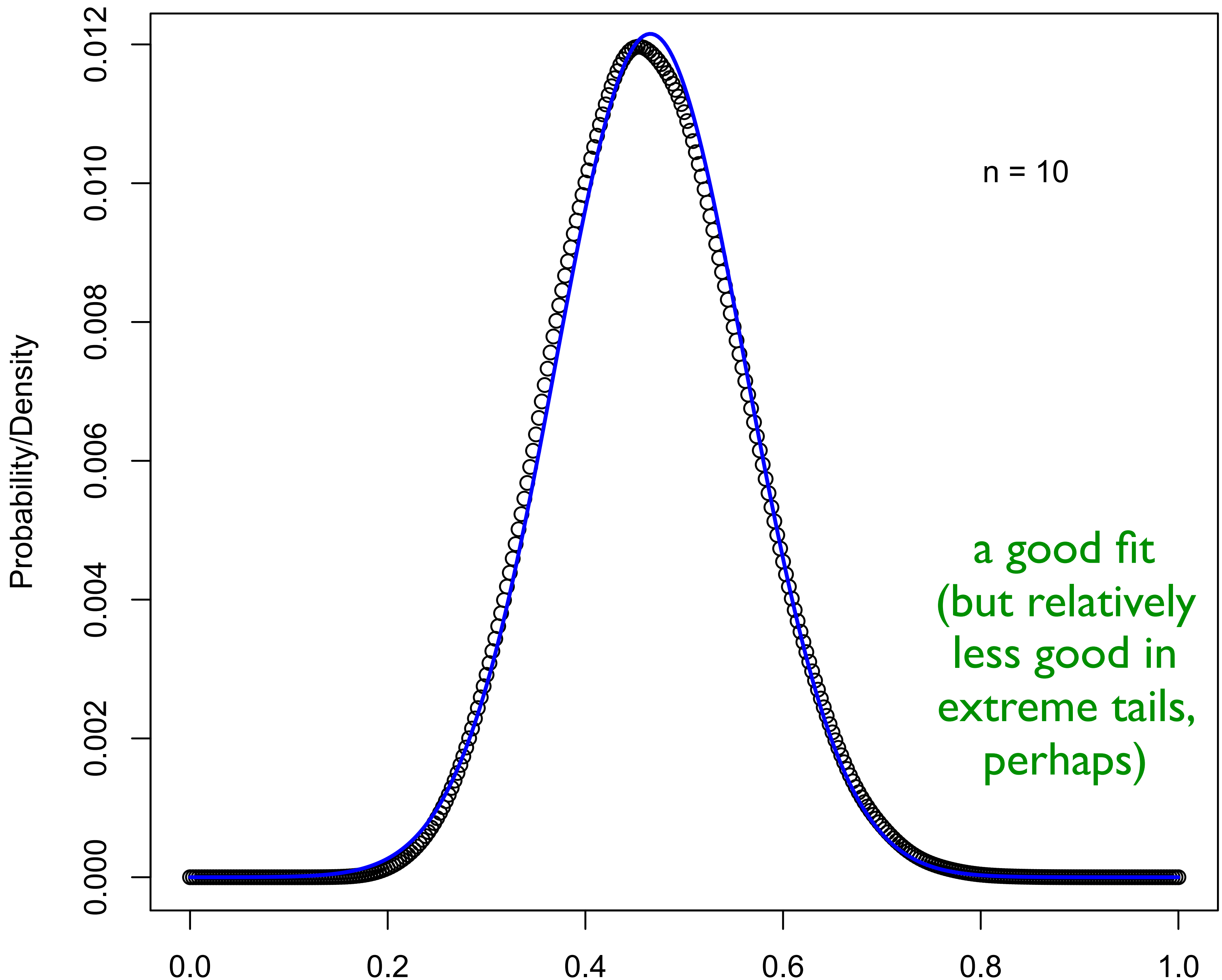
Note: on slide 5, showed sum of normals is *exactly* normal. Maybe not a surprise, given that sums of almost *anything* become *approximately* normal...


Probability/Density
x-bar
n = 1
n = 2
n = 3
n = 4

demo

CLT applies even to whacky distributions




Probability/Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.0
0.2
0.4
0.6
0.8
1.0
n = 10
a good fit (but relatively less good in extreme tails, perhaps)

CLT in the real world

CLT *also* holds under weaker assumptions than stated above, and is the reason many things appear normally distributed

Many quantities = sums of (roughly) independent random vars

Exam scores: sums of individual problems

People's heights: sum of many genetic & environmental factors

Measurements: sums of various small instrument errors

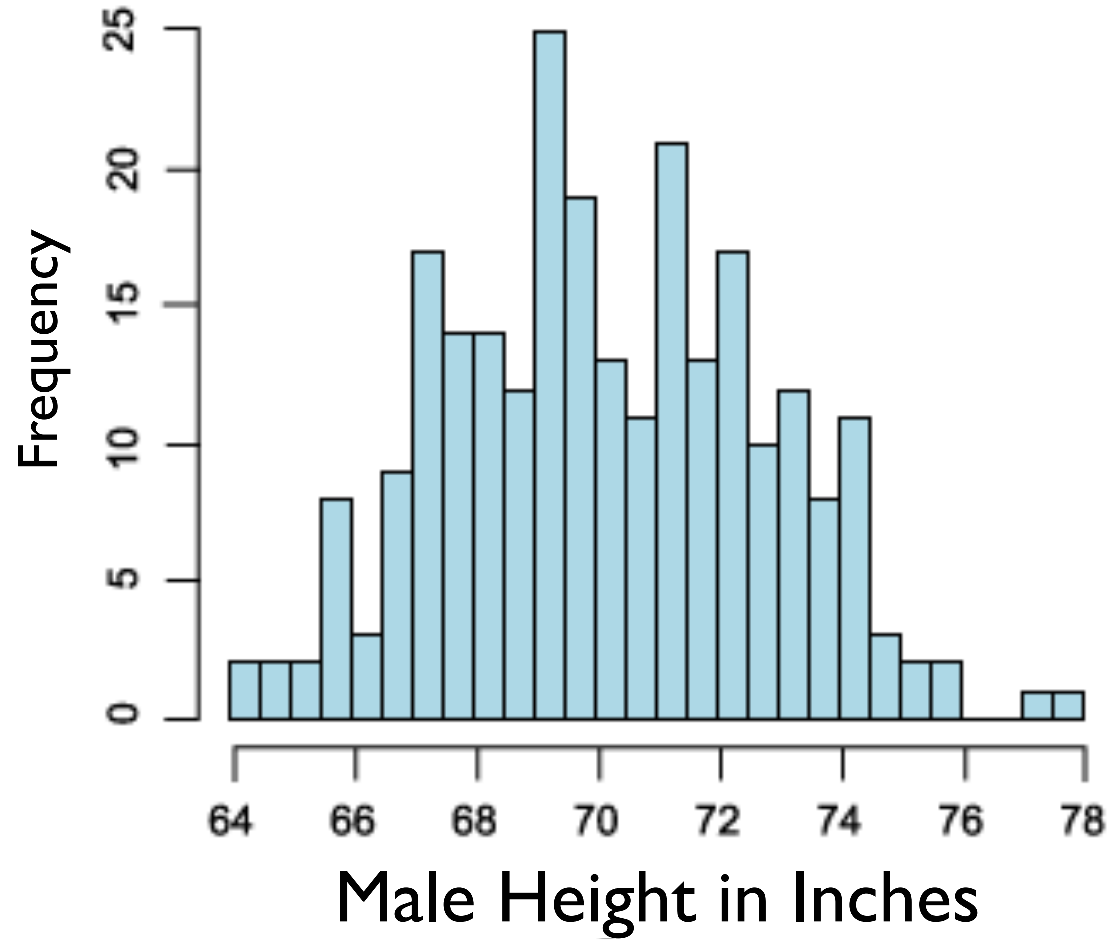
"Noise" in sci/engr applications: sums of random perturbations

...

Human height is approximately normal.

Why might that be true?

R.A. Fisher (1918) noted it would follow from CLT if height were the sum of many independent random effects, e.g. many genetic factors (plus some environmental ones like diet). I.e., suggested part of *mechanism* by looking at *shape* of the curve.



Let S_n = number of successes in n (indp.) trials (with prob. p).

$S_n \sim \text{Bin}(n,p)$ $E[S_n] = np$ $Var[S_n] = np(1-p)$

Poisson approx: good for n large, p small (np constant)

Normal approx: For large n, (p stays fixed):

$S_n \approx Y \sim N(E[S_n], Var[S_n]) = N(np, np(1-p))$

Rule of thumb: Normal approx "good" if $np(1-p) \geq 10$

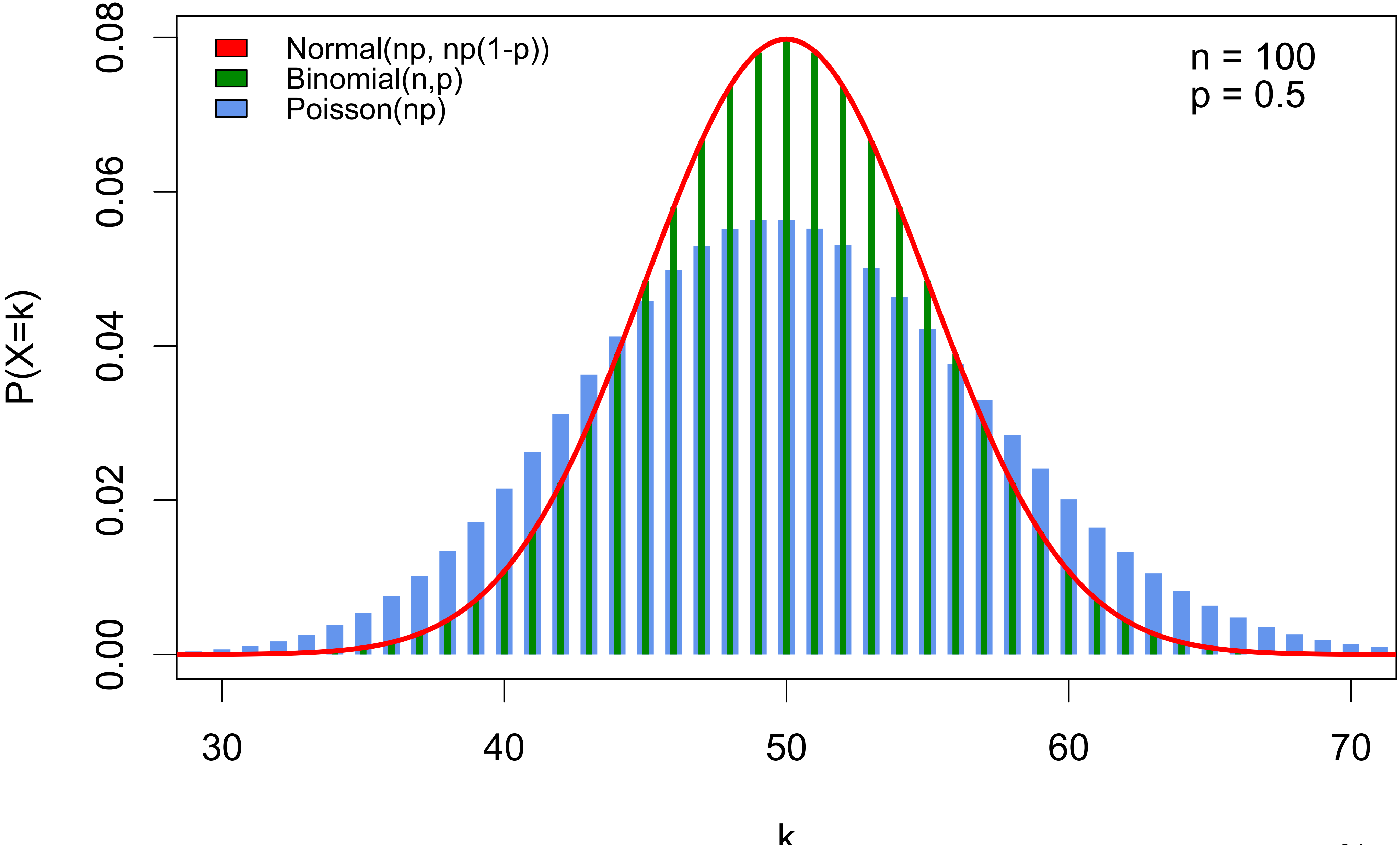
DeMoivre-Laplace Theorem:

As $n \rightarrow \infty$:

$$Pr\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \longrightarrow \Phi(b) - \Phi(a)$$

Equivalently:

$$Pr(a \leq S_n \leq b) \longrightarrow \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$


Normal(np, np(1-p))
Binomial(n,p)
Poisson(np)
n = 100
p = 0.5
P(X=k)
0.00
0.02
0.04
0.06
0.08
30
40
50
60
70
k

Ex: Fair coin flipped (independently) 40 times. Probability of 15 to 25 heads?

Exact (binomial) answer:

$$P_{\text{bin}}(15 \le X \le 25) = \sum_{i=15}^{25} \binom{40}{i} \left(\frac{1}{2}\right)^{40} \approx \boxed{0.9193}$$

Normal approximation:

$$\begin{aligned} P_{\text{norm}}(15 \le X \le 25) &= P\left(\frac{15-20}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} \le \frac{25-20}{\sqrt{10}}\right) \\ &\approx P\left(-1.58 \le \frac{X-20}{\sqrt{10}} \le 1.58\right) \\ &= \Phi(1.58) - \Phi(-1.58) \approx \boxed{0.8862} \end{aligned}$$

R Sidebar

```
> pbinom(25,40,.5) - pbinom(14,40,.5)
[1] 0.9193095

> pnorm(5/sqrt(10))- pnorm(-5/sqrt(10))
[1] 0.8861537
> 5/sqrt(10)
[1] 1.581139
> pnorm(1.58)- pnorm(-1.58)
[1] 0.8858931
```

SIDEBARS

I've included a few sidebar slides like this one (a) to show you how to do various calculations in R, (b) to check my own math, and (c) occasionally to show the (usually small) effect of some approximations. Feel free to ignore them unless you want to pick up some R tips.

Ex: Fair coin flipped (independently) 40 times. Probability of 20 or 21 heads?

Exact (binomial) answer:

$$P_{\text{bin}}(X = 20 \vee X = 21) = \left[\binom{40}{20} + \binom{40}{21}\right]\left(\frac{1}{2}\right)^{40} \approx \boxed{0.2448}$$

Normal approximation:

$$\begin{aligned} P_{\text{norm}}(20 \leq X \leq 21) &= P\left(\frac{20-20}{\sqrt{10}} \leq \frac{X-20}{\sqrt{10}} \leq \frac{21-20}{\sqrt{10}}\right) \\ &\approx P\left(0 \leq \frac{X-20}{\sqrt{10}} \leq 0.32\right) \\ &\approx \Phi(0.32) - \Phi(0.00) \approx \boxed{0.1241} \end{aligned}$$

Hmmm… A bit disappointing.

R Sidebar

```
> sum(dbinom(20:21,40,.5))
[1] 0.2447713
> pnorm(0)- pnorm(-1/sqrt(10))
[1] 0.1240852
> 1/sqrt(10)
[1] 0.3162278
> pnorm(.32)- pnorm(0)
[1] 0.1255158
```

Ex: Fair coin flipped (independently) 40 times. Probability of 20 heads?

Exact (binomial) answer:

$$P_{\text{bin}}(X = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx \boxed{0.1254}$$

Normal approximation:

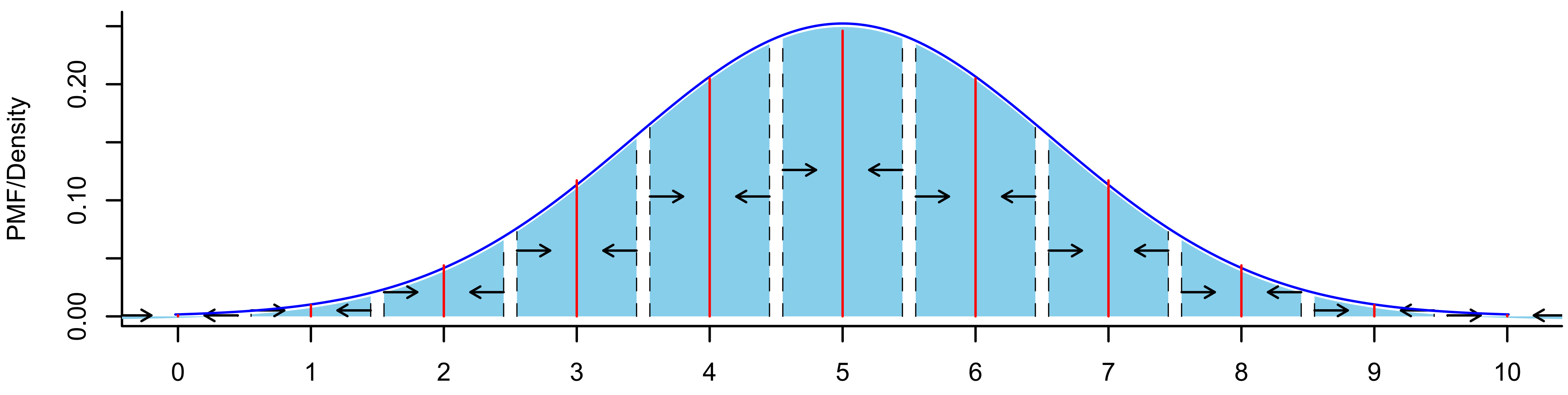
$$\begin{aligned} P_{\text{norm}}(20 \le X \le 20) &= P\left(\frac{20-20}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} \le \frac{20-20}{\sqrt{10}}\right) \\ &\approx P\left(0 \le \frac{X-20}{\sqrt{10}} \le 0\right) \\ &= \Phi(0.00) - \Phi(0.00) = \boxed{0.0000} \end{aligned}$$

Whoa! … Even more disappointing.

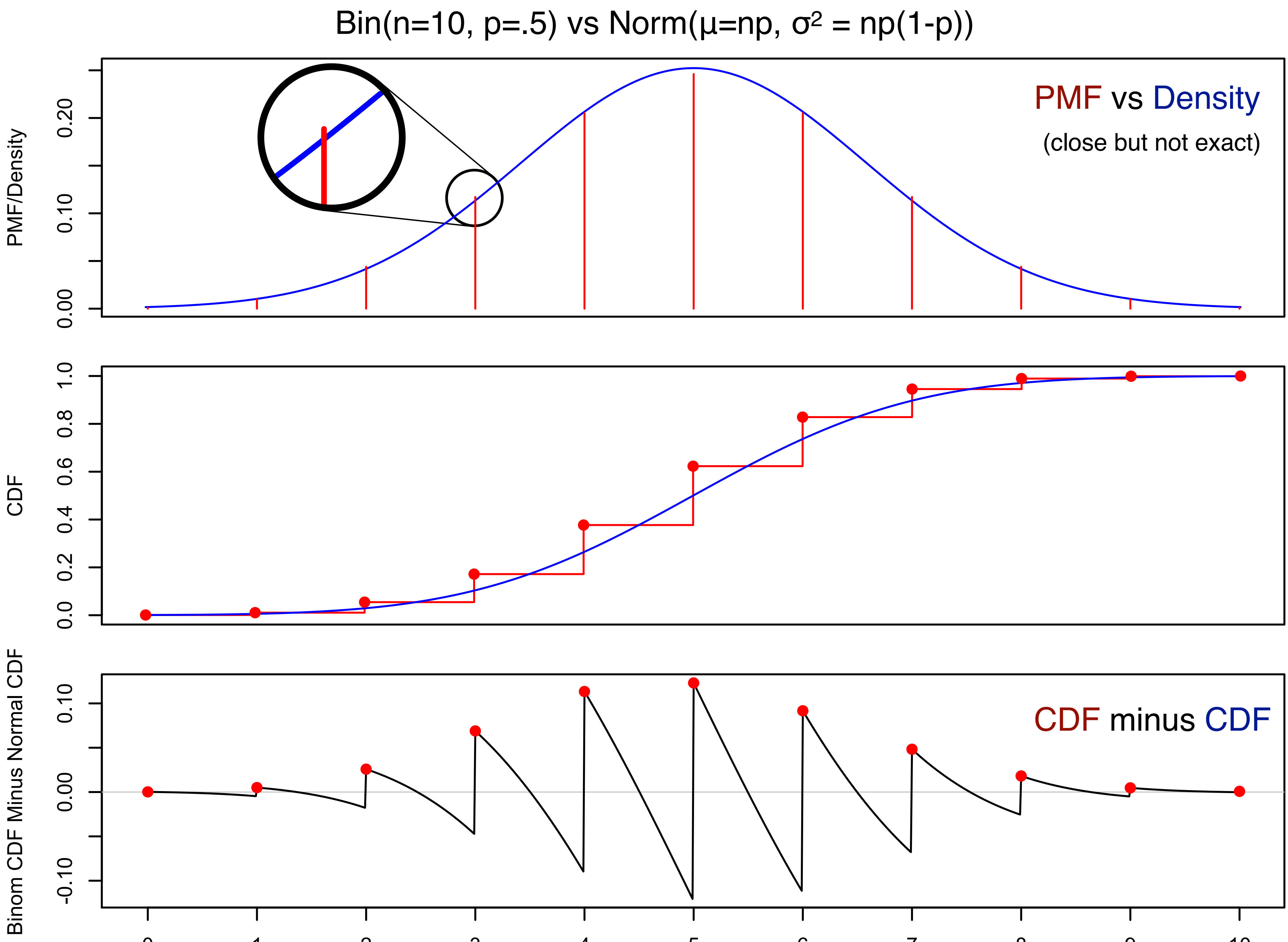
DeMoivre-Laplace and the "continuity correction"

Can we fix these anomalies? Yes! The "continuity correction":

Imagine *discretizing* the normal density by shifting probability mass at non-integer x to the nearest integer (i.e., "rounding" x). Then, probability of binom r.v. falling in the (*integer*) interval [a, ..., b], inclusive, is ≈ the probability of a normal r.v. with the same μ,σ^2 falling in the (*real*) interval [a-½ , b+½], even when a = b.



More: Feller, 1945 http://www.jstor.org/stable/10.2307/2236142



Bin(n=10, p=.5) vs Norm(μ=np, σ² = np(1-p))
PMF vs Density
(close but not exact)
PMF/Density
0.00
0.10
0.20
CDF
0.0
0.2
0.4
0.6
0.8
1.0
CDF minus CDF
Binom CDF Minus Normal CDF
-0.10
0.00
0.10
0
1
2
3
4
5
6
7
8
9
10

Ex: Fair coin flipped (independently) 40 times. Probability of 20 heads?

Exact (binomial) answer:

$$P_{\text{bin}}(X = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx \boxed{0.1254}$$

Normal approximation:

$$\begin{aligned} P_{\text{bin}}(X = 20) &\approx P_{\text{norm}}(19.5 \le X \le 20.5) \\ &= P_{\text{norm}}\left(\frac{19.5 - 20}{\sqrt{10}} \le \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \\ &\approx P_{\text{norm}}\left(-0.16 \le \frac{X - 20}{\sqrt{10}} < 0.16\right) \\ &= \Phi(0.16) - \Phi(-0.16) \approx \boxed{0.1272} \end{aligned}$$

{19.5 ≤ X ≤ 20.5} is the set of *reals* that round to the set of *integers* in {X = 20}

Ex: Fair coin flipped (independently) 40 times. Probability of 20 or 21 heads?

Exact (binomial) answer:

$$P_{\text{bin}}(X = 20 \vee X = 21) \;=\; \left[\binom{40}{20} + \binom{40}{21}\right]\left(\frac{1}{2}\right)^{40} \approx \boxed{0.2448}$$

Normal approximation:

$$\begin{aligned} P_{\text{bin}}(20 \le X < 22) &= P_{\text{norm}}(19.5 \le X \le 21.5) \\ &= P_{\text{norm}}\left(\frac{19.5-20}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} \le \frac{21.5-20}{\sqrt{10}}\right) \\ &\approx P_{\text{norm}}\left(-0.16 \le \frac{X-20}{\sqrt{10}} \le 0.47\right) \\ &\approx \Phi(0.47) - \Phi(-0.16) \approx \boxed{0.2452} \end{aligned}$$

{19.5 ≤ X < 21.5} is the set of *reals* that round to the set of *integers* in {20 ≤ X < 22}

One more note on continuity correction: Never wrong to use it, but it has the largest effect when the set of integers is small. Conversely, it's often omitted when the set is large.

R Sidebar

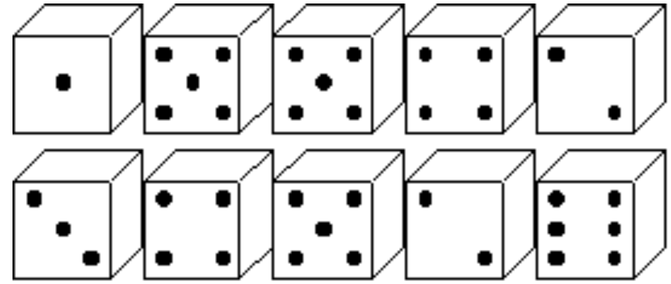
```
> pnorm(1.5/sqrt(10))- pnorm(-.5/sqrt(10))
[1] 0.2451883
> c(0.5,1.5)/sqrt(10)
[1] 0.1581139 0.4743416
> pnorm(0.47) - pnorm(-0.16)
[1] 0.244382
```

continuity correction is applicable beyond binomials

Roll 10 6-sided dice (independently)

X = total value of all 10 dice

Win if: X ≤ 25 or X ≥ 45



$$\text{E}[X] = \text{E}[\sum_{i=1}^{10} X_i] = 10\text{E}[X_1] = 10(7/2) = 35$$

$$\text{Var}[X] = \text{Var}[\sum_{i=1}^{10} X_i] = 10\text{Var}[X_1] = 10(35/12) = 350/12$$

$$P(\text{win}) = 1 - P\left(25.5 \leq X \leq 44.5\right) =$$

$$1 - P\left(\frac{25.5-35}{\sqrt{350/12}} \leq \frac{X-35}{\sqrt{350/12}} \leq \frac{44.5-35}{\sqrt{350/12}}\right)$$

$$\approx 2(1 - \Phi(1.76)) \approx 0.079$$

Poll of 100 randomly chosen voters finds that K of them favor proposition 666.

So: the *estimated proportion* in favor is $K/100 = q$

Suppose: the *true proportion* in favor is p.

Q. Give an upper bound on the probability that your estimate is off by > 10 percentage points, i.e., the probability of $|q - p| > 0.1$

A. $K = X_1 + \ldots + X_{100}$, where X_i are Bernoulli(p), so by CLT:

$K \approx$ normal with mean $100p$ and variance $100p(1-p)$; or:

$q \approx$ normal with mean p and variance $\sigma^2 = p(1-p)/100$

Letting $Z = (q-p)/\sigma$ (a standardized r.v.), then $|q - p| > 0.1 \Leftrightarrow |Z| > 0.1/\sigma$

By symmetry of the normal

$P_{Ber}(\ |q - p| > 0.1\ ) \approx 2\ P_{norm}(\ Z > 0.1/\sigma\ ) = 2\ (1 - \Phi(0.1/\sigma))$

Unfortunately, p & σ are unknown, but $\sigma^2 = p(1-p)/100$ is maximized when $p = 1/2$, so $\sigma^2 \leq 1/400$, i.e. $\sigma \leq 1/20$, hence

$2\ (1 - \Phi(0.1/\sigma)) \leq 2(1-\Phi(2)) \approx$ **0.046**

Exercise: How much smaller can σ be if p ≠ 1/2?

I.e., less than a 5% chance of an error as large as 10 percentage points.

Distribution of X + Y: summations, integrals (or MGF)

Distribution of X + Y ≠ distribution X or Y in general

Distribution of X + Y is normal if X and Y are normal (*)

(ditto for a few other special distributions)

Sums generally don't "converge," but averages do:

Weak Law of Large Numbers

Strong Law of Large Numbers

Most surprisingly, averages often converge to the *same* distribution:

the Central Limit Theorem says sample mean → normal

[Note that (*) essentially a prerequisite, and that (*) is exact, whereas CLT is approximate]